*Version 1.2*

# Definition of Data Quality

Issued: 05 Jan 06

*Census Bureau Principle*

Authored for:

Census Bureau Methodology & Standards Council
Al Tupek, Chair

USCENSUSBUREAU

*Helping You Make Informed Decisions*

# Document Management & Control [1]

| Version | Issue Date | Approval | Description |
|---|---|---|---|
| 1.0 | 22 Nov 05 | M&S Council | Initial Release |
| 1.1 | 13 Dec 05 | Configuration Mgr. | Re-Edit |
| 1.2 | 05 Jan 06 | Configuration Mgr. | Reformat |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

[1] **The most current version of this document is maintained on the Census Bureau Intranet and may be accessed from the Quality Management Repository.**

# Census Bureau Principle: Definition of Data Quality

The Census Bureau defines quality as "fitness for use." We are guided by the needs of our customers to ensure our data products are fit for their use. Our customers include all branches and levels of the Federal government, state and local governments, and the public. When planning each data product we identify our primary customers and their uses for the data. We provide this definition of data quality to help our customers and employees understand the foundation of our data quality principles and standards.

We characterize fitness for use in terms of three attributes that the Office of Management and Budget (OMB) expects of the information provided by our data products:

**Utility** - refers to the usefulness of the information to its intended users.

**Objectivity** - refers to whether information is accurate, reliable, and unbiased, and is presented in an accurate, clear, and unbiased manner.

**Integrity** - refers to the security or protection of information from unauthorized access or revision.

To help apply limited resources to best achieve fitness for use, the Census Bureau further defines utility, objectivity, and integrity in terms of six dimensions of data quality: relevance, accuracy, timeliness, accessibility, interpretability, and transparency. Thinking of data quality in these dimensions helps the Census Bureau to analyze the trade-offs between the sometimes conflicting dimensions, to meet our customers' needs, and ultimately, to satisfy our customers.

**Relevance** refers to the degree to which our data products provide information that meets our customers' needs.

**Accuracy** refers to the difference between an estimate and its true value. We characterize the difference in terms of systematic (bias) and random (variance) errors.

**Timeliness** refers to the length of time between the reference period of the information and when we deliver the data product to our customers.

**Accessibility** refers to the ease with which customers can identify, obtain, and use the information in our data products.

**Interpretability** refers to the availability of documentation to aid customers in understanding and using our data products. This documentation typically includes a presentation of: the underlying concepts; definitions; descriptions of the methods used to collect, process, and analyze the data; and a discussion of the limitations imposed by the methods used.

**Transparency** refers to the existence of evidence that customers can use to assess the accuracy of the data product. This evidence would include information on the assumptions, methods, and results such that a qualified third party could reproduce the information, within the constraints of protecting respondent confidentiality and privacy.

In planning, developing, and delivering a data product, these dimensions of data quality may come into conflict. When this occurs, the Census Bureau analyzes all the dimensions to achieve a suitable balance between them. For example, timeliness and accuracy may present conflicting priorities. To achieve the best accuracy might require a longer data collection period to follow up nonrespondents and obtain interviews. But if obtaining near-perfect response would delay the data product until it was too late to inform the customer's decision, then a trade-off might be needed between accuracy and timeliness to achieve fitness for use. Similarly, unexpected circumstances in collecting or processing the data might require the Census Bureau to re-evaluate the dimensions of quality and redirect resources aimed at satisfying one dimension to focus on another, in the interest of achieving fitness for use.

## Approved by the Census Bureau Methodology and Standards Council:

*Al R. Tupek* — *Signed November 22, 2005*

Alan R. Tupek — Date
Chair, Methodology and Standards Council

*Jimmie B. Scott* — *Signed November 22, 2005*

Jimmie B. Scott — Date
Acting Chief, Demographic Statistical Methods Division

*R Singh* — *Signed November 22, 2005*

Rajendra P. Singh — Date
Chief, Decennial Statistical Studies Division

*Howard Hogan* — *Signed November 22, 2005*

Howard Hogan — Date
Chief, Economic Statistical Methods and Programming Division

*Ruth Ann Killion* — *Signed November 22, 2005*

Ruth Ann Killion — Date
Chief, Planning, Research, and Evaluation Division

*Tommy Wright* — *Signed November 22, 2005*

Tommy Wright — Date
Chief, Statistical Research Division